



UNIVERSITY OF  
CAMBRIDGE



Cambridge  
Data Champions

# Research Data Management – an introduction

Laura Kennedy – [lhk24@cam.ac.uk](mailto:lhk24@cam.ac.uk), [@laurahkennedy](https://twitter.com/laurahkennedy)

Ben Jarman – [bmhj2@cam.ac.uk](mailto:bmhj2@cam.ac.uk), [@bjarman\\_](https://twitter.com/bjarman)

# What counts as 'data'?

- Many kinds...

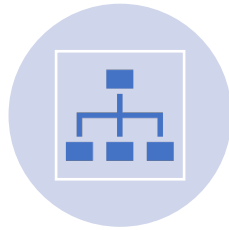
- Raw data (e.g. fieldnotes, interview recordings, etc.)
- Processed data
- Analysed data
- Photos
- Videos
- Surveys and interviews
- Newspapers (or any other texts)
- Field notes
- Maps
- Notebooks
- Spreadsheets
- Etc.



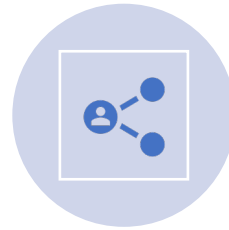
# What is 'research data management'?



PLANNING



CREATING



ORGANISING



SHARING

# Aims

Reasons to  
think about  
RDM

Key concepts  
and good  
practices

Pointers to  
resources

Time to discuss  
application to  
your work



# Overview of subject matter

## Good practices (Ben)

Backup and file  
sharing

Organising data

Personal and  
sensitive data

## Improving research (Laura)

Open data,  
materials, and code

Data management  
plans

Further resources

## Over to you...

Questions

Practical queries?



## Part 1:

# Data backup and file sharing strategies



# Avoiding catastrophe



Credit: Peter Murray-Rust,  
<http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan/>, August 2011, CC-BY

Schofield's  
Second Law of  
computing:

“Data doesn't really  
exist unless you have at  
least two copies of it”



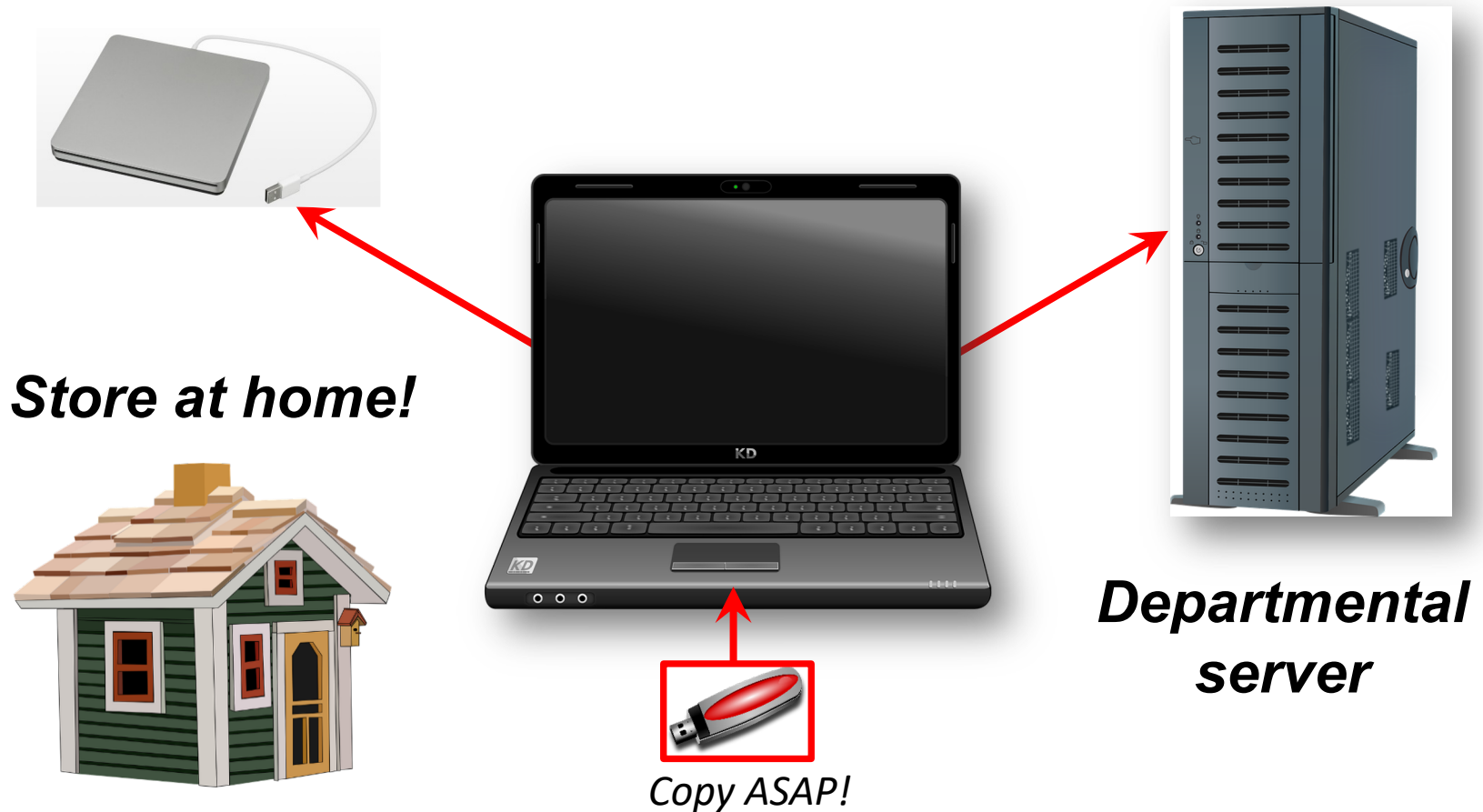
# What would happen if...

- your laptop got stolen?
- your department had a fire?
- you lost your USB stick?
- your external hard disk stopped working?
- data from your Dropbox/Google Drive/OneDrive account disappeared

The screenshot shows the BBC News website. The main headline is "Swansea University fire: Emergency services tackle blaze at Bay campus", dated 12 August 2020. Below the headline is a video player showing thick black smoke rising from a building. A caption below the video states: "The fire occurred at the university's school of engineering". The article text below the video reads: "A fire has broken out at Swansea University's Bay campus. Mid and West Wales Fire and Rescue Service (MAWWFRS) sent five appliances to the scene, as well as a water bowser and aerial appliance. The university **tweeted to say the fire had broken out** in the engineering department and officials were co-operating with fire officers. There were no reports of anyone inside the building, said MAWWFRS."

On the right side of the page, there are "Top Stories" including "Vaccine minister 'confident' of supplies" (dated 24 minutes ago) and "Ministers to discuss new UK travel restrictions" (dated 2 hours ago). Below these are "Features" including "EU to tighten vaccine exports amid AstraZeneca row" (dated 52 minutes ago) and "Five days that shaped the outbreak" (with an accompanying image of medical professionals).

# Backup – as easy as 3-2-1...



Free software to manage backups (there are many, many alternatives):

<http://www.2brightsparks.com/download-syncbackfree.html>

# File sharing

- E-mail
- Website (Moodle, Confluence, Basecamp etc.)
  - via FTP/SFTP?
- University of Cambridge cloud services
  - OneDrive
  - Dropbox
  - GSuite
- Other cloud services (e.g. Google Drive)
  - Caution - do not use cloud storage for restricted data
  - i.e. “documents containing sensitive or personal information or that were intended only for the use of a named person”

# File sharing

E-mail	<ul style="list-style-type: none"><li>• Think before you send!</li></ul>
Website (Moodle, blogs, etc.)	<ul style="list-style-type: none"><li>• via FTP/SFTP?</li><li>• How uploaded?</li><li>• Access management?</li></ul>
University of Cambridge cloud services	<ul style="list-style-type: none"><li>• OneDrive</li><li>• Dropbox</li><li>• GSuite</li></ul>
Other cloud services (e.g. Google Drive)	<ul style="list-style-type: none"><li>• Do <u>not</u> use cloud storage for restricted data</li><li>• i.e. “documents containing sensitive or personal information or that were intended only for the use of a named person”</li></ul>



# OneDrive - University of Cambridge



## Set up your 1TB OneDrive for Business personal filestore\*




- Sync content across your devices
- Access files from virtually anywhere
- Share files with colleagues

*\* Applies to those qualifying under the University's EES agreement*

University OneDrive filespaces use the OneDrive for Business solution which gives you 1TB of cloud-based file storage space for your work and personal files. Find out how to access your online filestore, and start syncing your local files.

<http://www.uis.cam.ac.uk/ees/onedrive>  
Questions: [service-desk@uis.cam.ac.uk](mailto:service-desk@uis.cam.ac.uk)

# Cloud services (using Raven to sign up)

			
Integration with Office365	Yes	No	No
Space included	5TB	Unlimited	Unlimited
Price (per annum)	Free	Free	£84 p.a.
File history/recovery	90 days	30 days	Unlimited
Backups	In the Microsoft Cloud	Managed by Google	Managed by Dropbox
Support available	Webpages from UIS	Unsupported	Unsupported
Where are files stored?	UK data centres	Anywhere in the world	EU data centres
Live editing of files	Yes – dependent on browser	Yes	Only with an Office 365 licence
What types of data can be shared?	Public, Cambridge only, or confidential data	Public data	Public, Cambridge only, or confidential data

Full details: <https://help.uis.cam.ac.uk/service/storage/storage-overview>



Part 2:

# Data organisation



**Cambridge  
Data Champions**

# Data organisation

Which example is better and why?

## Example A

### Documents library

PhD data

- 25July
- Documents
- Experiments
- Experiments2
- From desktop
- Important
- Other
- PhD
- Talks
- Experiment 1
- Lab meeting FINAL
- Meeting notes
- My talk
- Paper submission
- PhD revised

## Example B

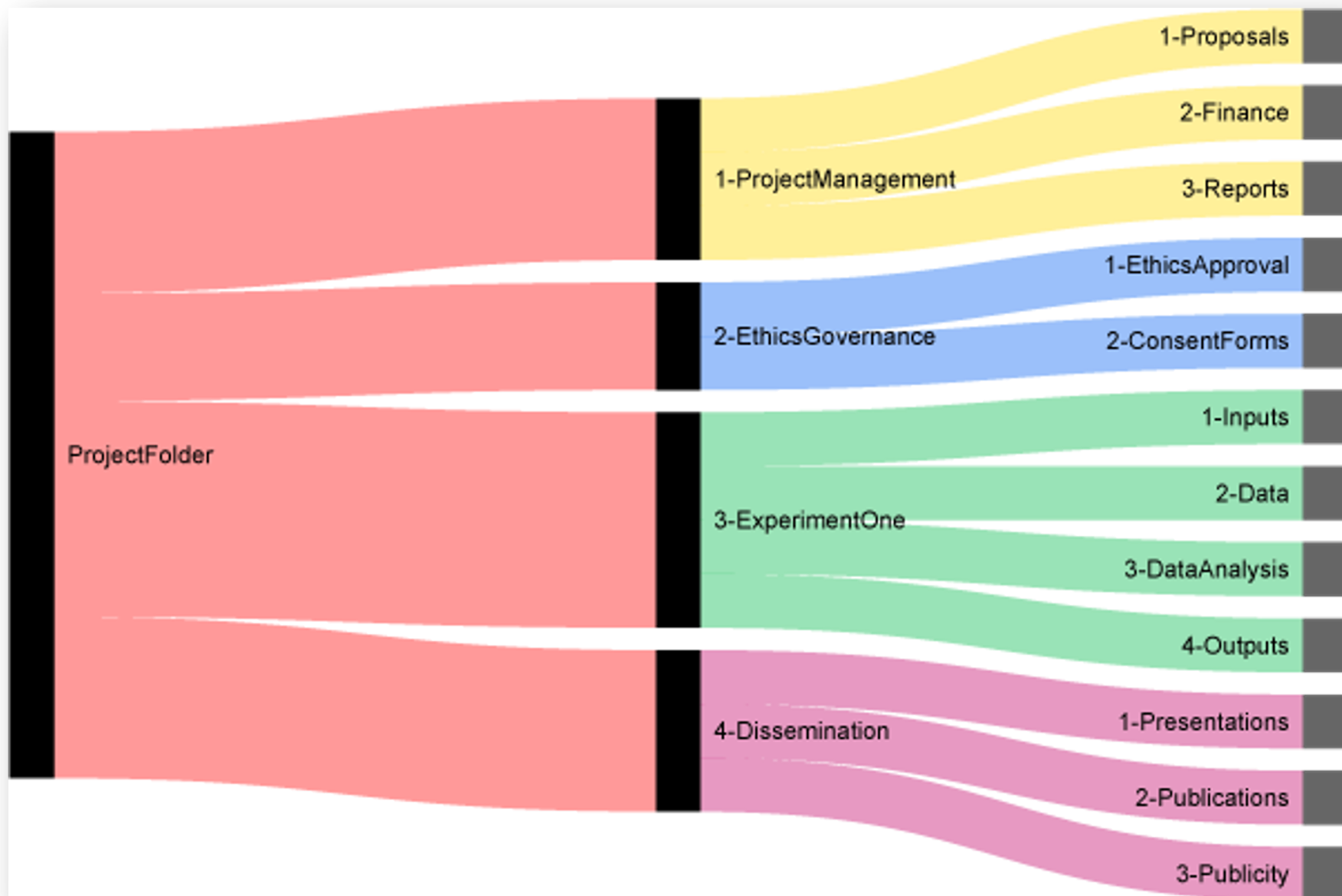
### Documents library

PhD data2

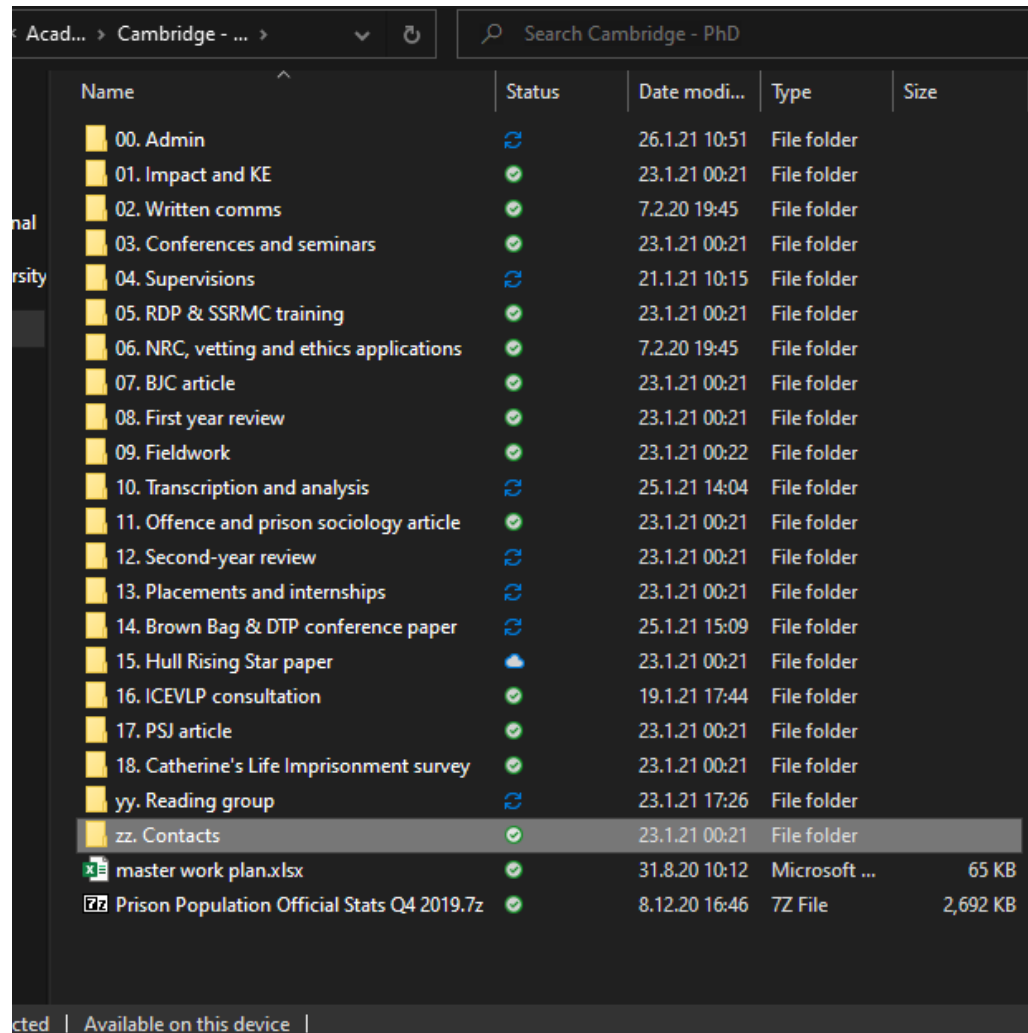
- Conferences
- Downloaded publications
- Experimental data
- Financial documents
- PhD thesis
- Presentations
- Protocols
- Reagents
- Reports
- Training

# Data organisation:

- Consistent
- Meaningful to you (and any colleagues you're working with)
- Allow you to find files easily
- Would you be able to easily find your data?

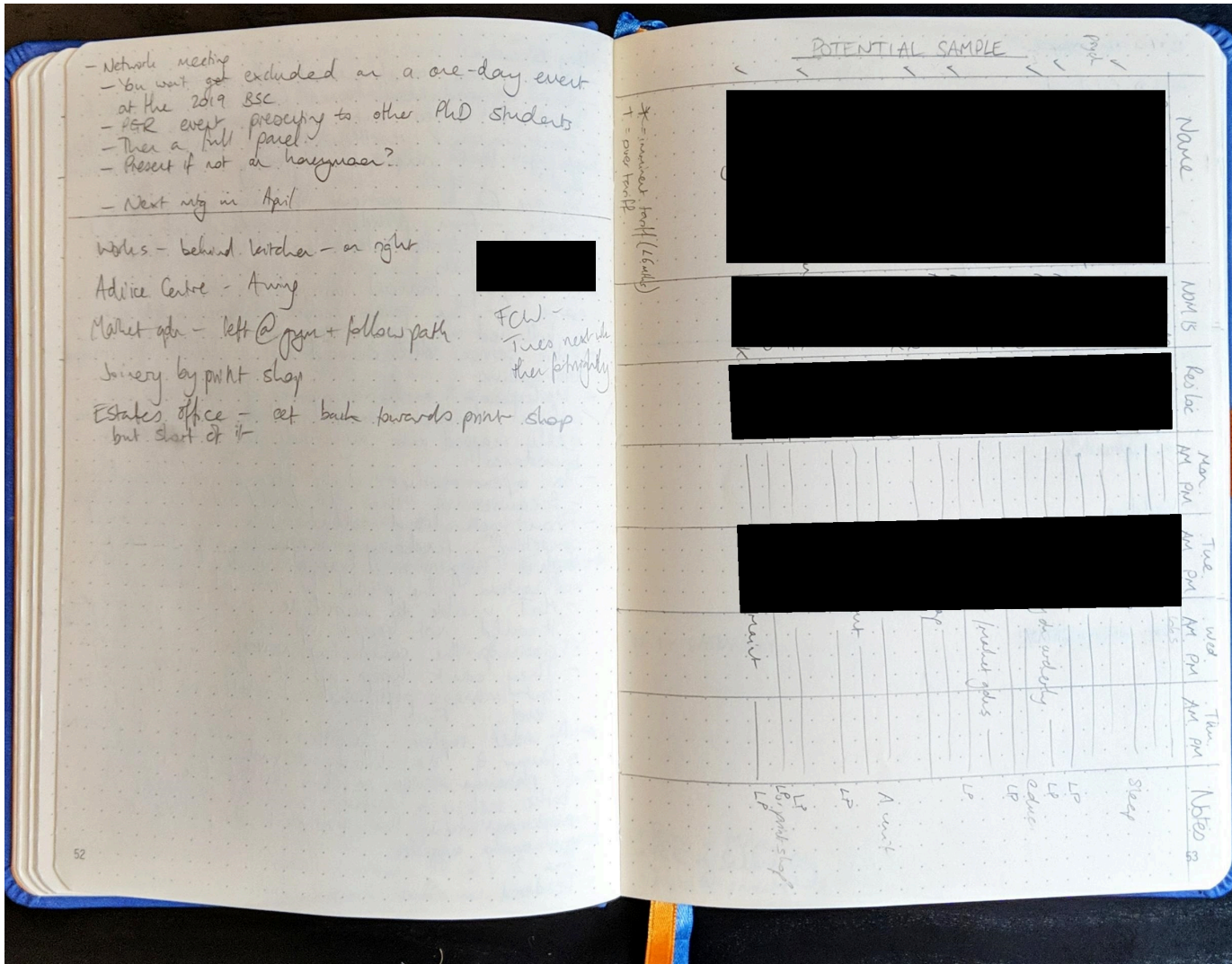


# PhD files



















Name	Status	Date modified	Type	Size
00. Admin	🔄	26.1.21 10:51	File folder	
01. Impact and KE	✅	23.1.21 00:21	File folder	
02. Written comms	✅	7.2.20 19:45	File folder	
03. Conferences and seminars	✅	23.1.21 00:21	File folder	
04. Supervisions	🔄	21.1.21 10:15	File folder	
05. RDP & SSRMC training	✅	23.1.21 00:21	File folder	
06. NRC, vetting and ethics applications	✅	7.2.20 19:45	File folder	
07. BJC article	✅	23.1.21 00:21	File folder	
08. First year review	✅	23.1.21 00:21	File folder	
09. Fieldwork	✅	23.1.21 00:22	File folder	
10. Transcription and analysis	🔄	25.1.21 14:04	File folder	
11. Offence and prison sociology article	✅	23.1.21 00:21	File folder	
12. Second-year review	🔄	23.1.21 00:21	File folder	
13. Placements and internships	🔄	23.1.21 00:21	File folder	
14. Brown Bag & DTP conference paper	🔄	25.1.21 15:09	File folder	
15. Hull Rising Star paper	☁	23.1.21 00:21	File folder	
16. ICEVLP consultation	✅	19.1.21 17:44	File folder	
17. PSJ article	✅	23.1.21 00:21	File folder	
18. Catherine's Life Imprisonment survey	✅	23.1.21 00:21	File folder	
yy. Reading group	🔄	23.1.21 17:26	File folder	
zz. Contacts	✅	23.1.21 00:21	File folder	
master work plan.xlsx	✅	31.8.20 10:12	Microsoft ...	65 KB
Prison Population Official Stats Q4 2019.7z	✅	8.12.20 16:46	7Z File	2,692 KB

# What data do you need to organise, and how?





# Interviews

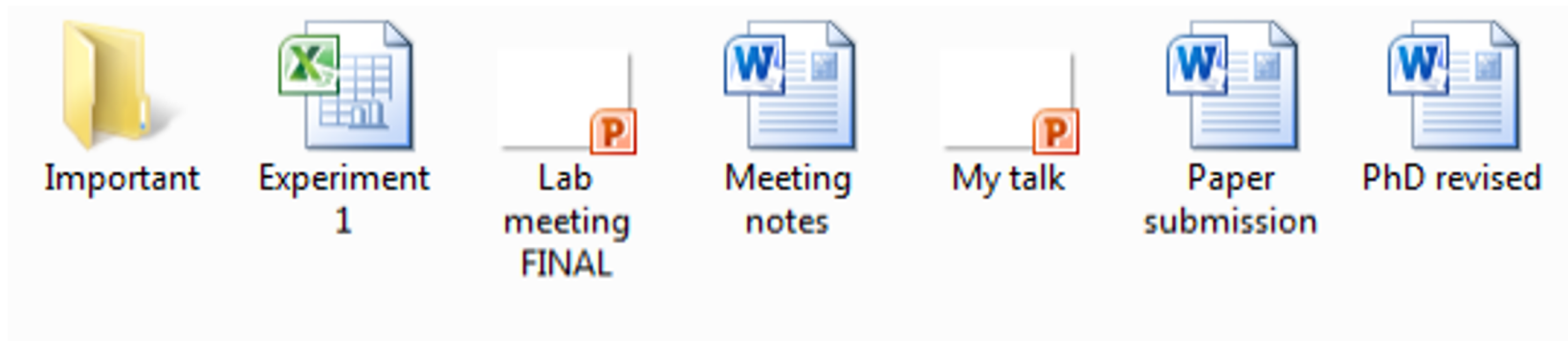
Name	Date modified	Type
 190917	10.7.20 11:04	OpenDocu
 190925	15.7.20 20:32	OpenDocu
 190926	2.9.20 18:41	OpenDocu
 190926	2.9.20 18:28	OTR File
 191001	11.8.20 11:07	OpenDocu
 191001	9.6.20 17:35	Microsoft
 191001	6.5.20 10:25	OTR File
 191002	15.7.20 20:15	OpenDocu
 191002	15.7.20 20:15	OpenDocu
 191003	9.6.20 17:35	OpenDocu
 191003	21.5.20 16:43	OTR File
 191008	9.6.20 17:36	Microsoft
 191009	10.7.20 11:01	OpenDocu
 191009	8.9.20 19:50	OpenDocu
 191009	8.9.20 19:40	OTR File
 191010	3.9.20 19:11	OpenDocu

# File naming conventions – do they matter?



Copyright: <http://10pm.com/>

# File naming conventions – do they matter?

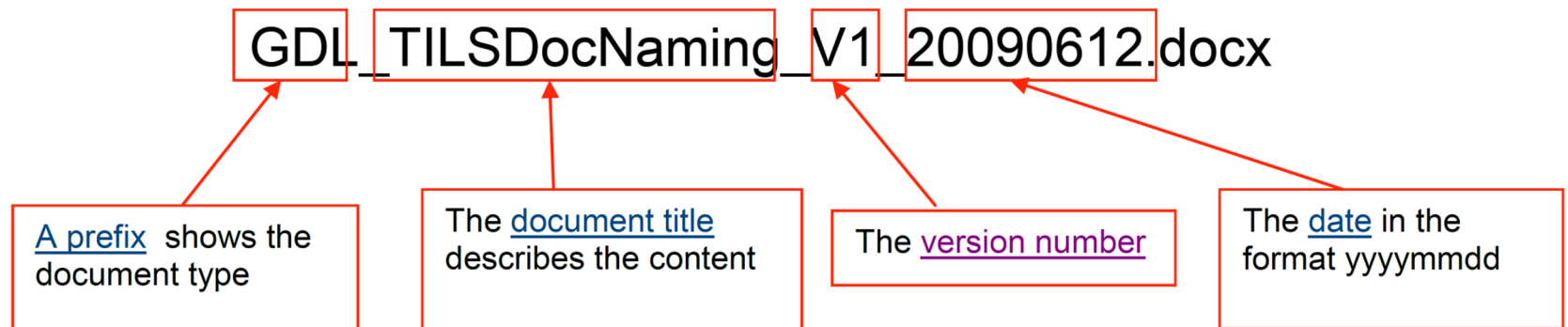


In 3 years time would you know what these are?

# File naming conventions











## TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:



[http://www.data.cam.ac.uk/files/gdl\\_tilsdocnaming\\_v1\\_20090612.pdf](http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf)












# File naming conventions

- ▶ 20160706\_Presentations
- ▶ 20160726\_RDMworkshopForLibrarians\_MillLane
- ▶ 20160802\_Altmetrics
- ▶ 20160804\_Bibliometrics
- ▼ 20160810\_RDMworkshopForLibrarians\_MilsteinRoom
  -  GDL\_DMP\_V8\_2016511.docx
  -  GDL\_DMPForLibrarians\_V9\_20160726docx.docx
  -  GDL\_ExampleDMP\_V9\_20160511.docx
  -  LST\_RDMforLibsAttendees\_V1\_20160808.pdf
  -  MEM\_RDMforLibs2Feedback\_V1\_20160811.docx
  -  PRE\_DataLossScenarios\_V5\_20160726.docx
  -  PRE\_RDMforLibsSigns\_V1\_20160808.pptx
  -  PRE\_RDMWorkshopForLibrarians\_V3\_20160726.pptx
  -  PRE\_RDMWorkshopForLibrarians\_V4\_20160810.pdf
  -  PRE\_RDMWorkshopForLibrarians\_V4\_20160810.pptx
- ▶ 20160914\_RDMWorkshopforGLS
- ▶ 20161004\_PhDTraining

[http://www.data.cam.ac.uk/files/gdl\\_tilsdocnaming\\_v1\\_20090612.pdf](http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf)

<< Cambri... > 07. BJC article

Search 07. BJC article

Name	Status	Date modified
 190812 only one way to swim - author's original version.docx	✓	20.4.20 12:1
 190812 only one way to swim - author's original version.pdf	✓	20.4.20 12:1
 200416 only one way to swim - accepted manuscript.docx	✓	22.4.20 12:1
 200416 only one way to swim - accepted manuscript.pdf	✓ 	8.4.20 22:46
 200420 copyright agreement confirmation.pdf	✓	20.4.20 12:3
 200420 copyright agreement.pdf	✓	20.4.20 12:3
 200425 corrected proofs.7z	✓	25.4.20 21:3
 drafts.7z	✓	20.4.20 12:1
 section word counts.xlsx	✓	3.4.20 16:52
 SSRN-id3585421.pdf	✓	25.4.20 22:0

<< Data ... > RDMTraining >

Search RDMTraining

Name

Status

Date modified



PRE\_OtherPresentationsForInspiration



18.1.21 12:22



EML\_20201103\_ProposalforWorkshops.docx



4.11.20 09:24



EML\_20201203\_AbstractToCaroline.eml



18.1.21 10:17



MTG\_20201112\_LauraCarolinePlanningZoomNotes.docx



18.1.21 10:38



MTG\_20210118\_LauraBenPlanningZoomNotes.docx



18.1.21 12:17



PRE\_20210118\_MPhilRDMTrainingSlides\_v1.pptx



26.1.21 11:50



Cambridge  
Data Champions

# Data Tools

Data  
cleaning

- e.g. OpenRefine

Data  
visualisation

- e.g. Qlik, Tableau

Electronic  
notebooks

- e.g. OneNote, Evernote

Text-based  
analysis

- e.g. Nvivo, Atlas.ti

## Schofield's Third Law..

- the easier it becomes for you to access your data, the easier it becomes for someone else to access your data

## So...

- Check where tools store data
- Take care over access
- Use strong passwords, password managers, and 2FA wherever possible





## Part 3:

# Managing personal and sensitive data



# What is personal data?



“records or other information that, on their own, or linked with other data or information in the possession of the data controller, can reveal the identity of an actual living person”



*Data Protection Act (1998)*

# What is sensitive data?

Personal  
data +  
information  
about

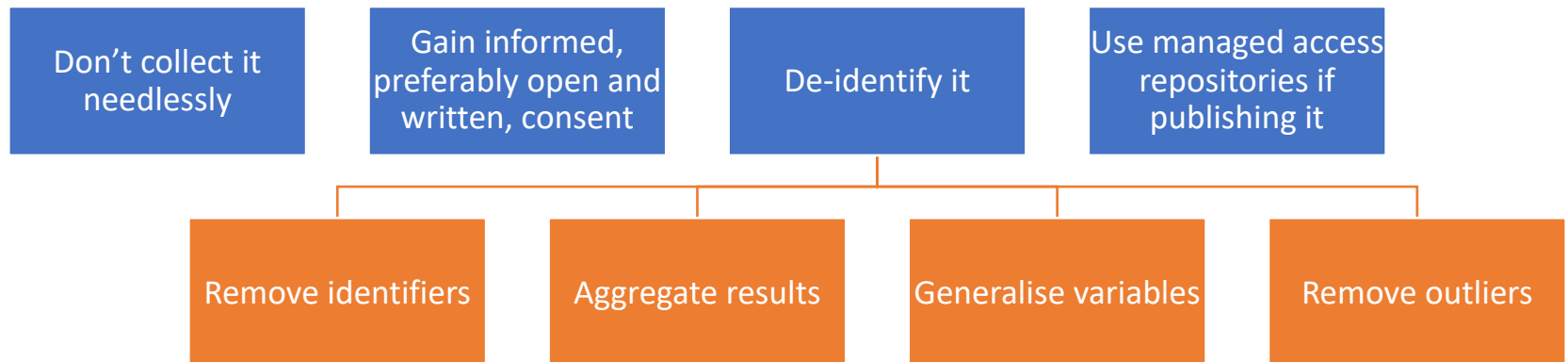
- racial or ethnic origin
- political opinions
- religious beliefs / other beliefs
- trade union membership
- physical or mental health or condition
- sexual life
- Offences/court proceedings...

# What is sensitive data?

Personal data PLUS information about:

- racial or ethnic origin
- political opinions
- religious beliefs / other beliefs
- trade union membership
- physical or mental health or condition
- sexual life
- offences/court proceedings...

# How to handle personal data?



# Resources for working with personal/sensitive data

University  
Research  
Integrity website

BSC Code of  
Ethics

ESRC guidelines

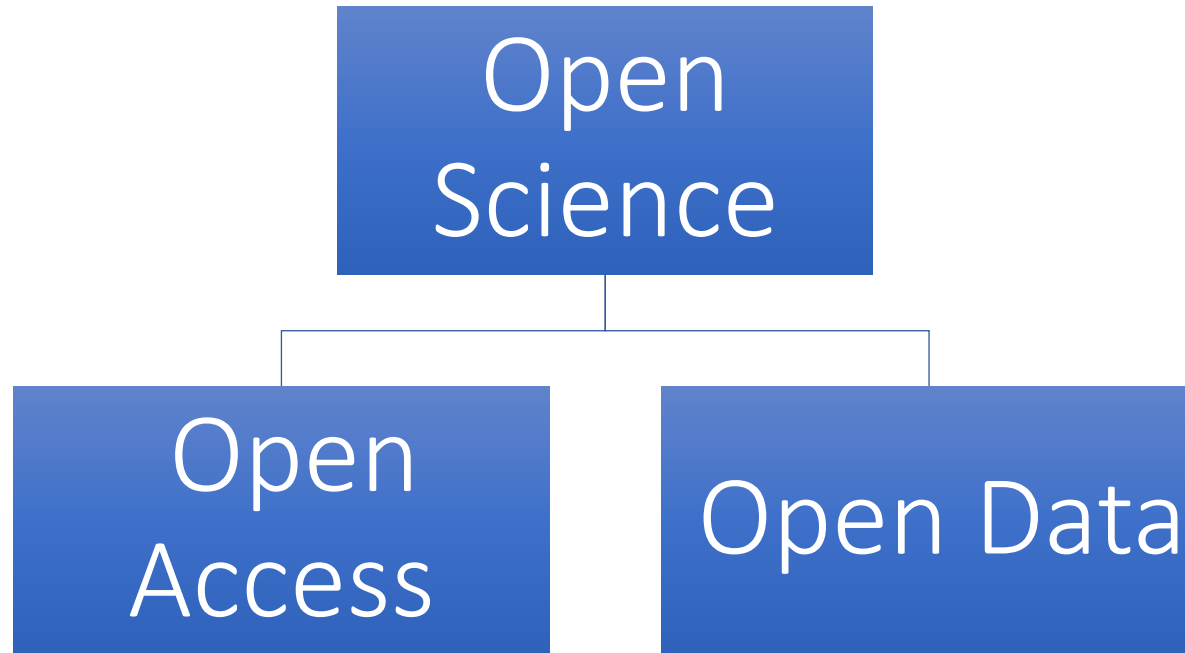
Cambridge Data  
Champions  
pages

Us (Ben and  
Laura!)



## Part 4: Open data, materials, and code



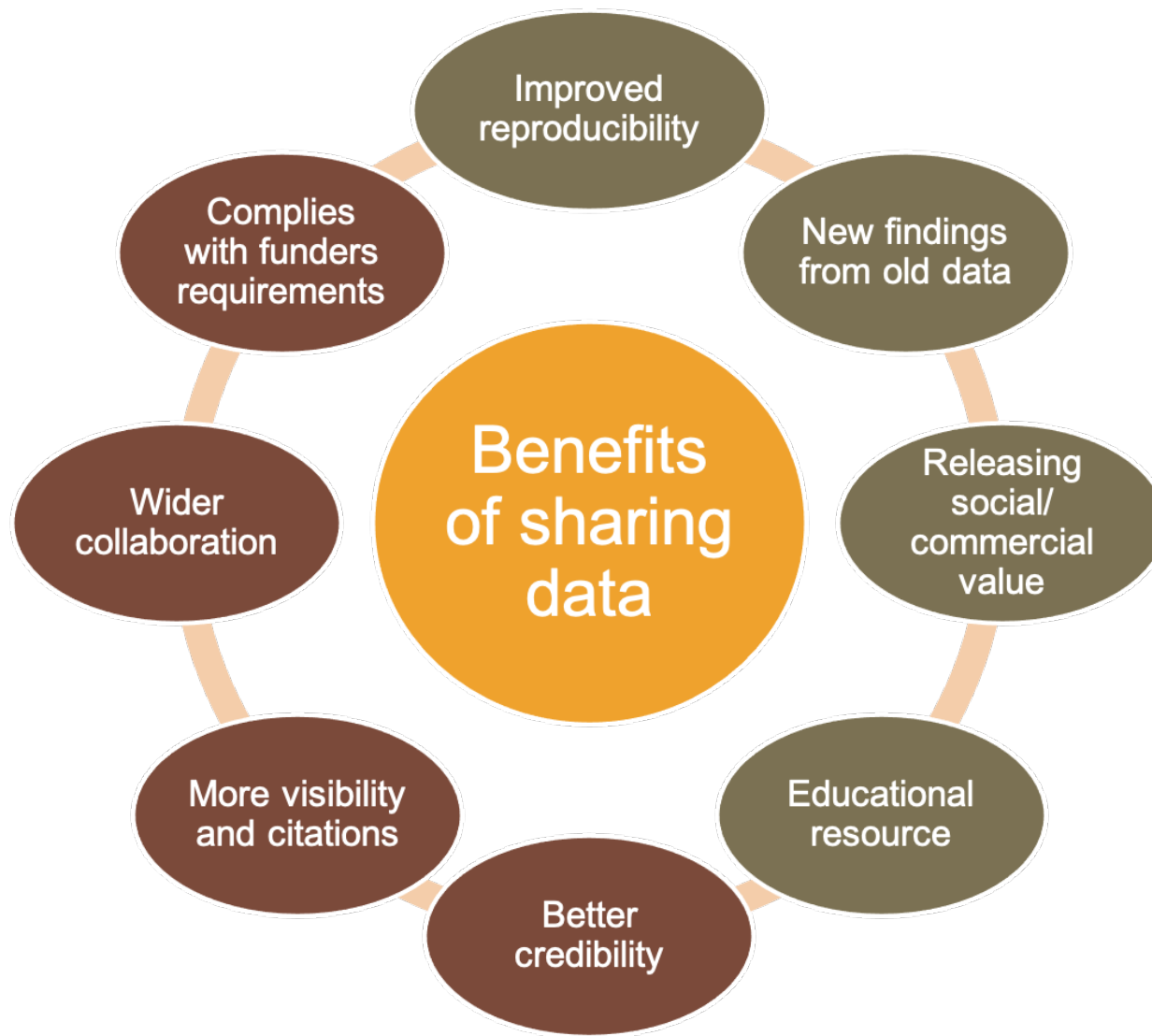


Crüwell et al. (2019), Seven easy steps to Open Science: an annotated reading list

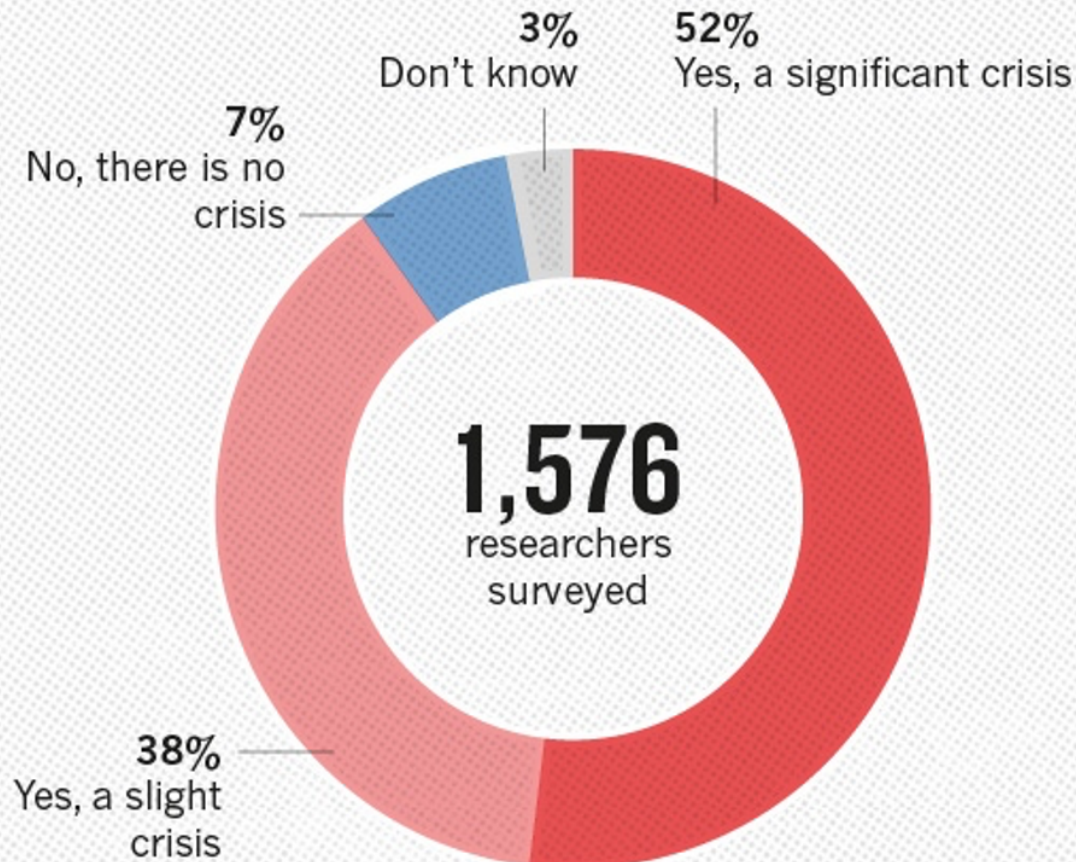
Tennant et al. (2016), The academic, economic and societal impacts of Open Access: an evidence-based review

Klein et al. (2018), A practical guide for transparency in psychological science





## IS THERE A REPRODUCIBILITY CRISIS?



©nature

Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a

*Annual Review of Criminology*

## Replication in Criminology and the Social Sciences

William Alex Pridemore,<sup>1</sup> Matthew C. Makel,<sup>2</sup>  
and Jonathan A. Plucker<sup>3</sup>

"[M]uch research in criminology is translatable to policy and practice, which can be expensive and have unintended harmful consequences. Thus, **the trustworthiness of criminology's scientific record is not only fundamental to its sustainability as a field but also integral to its relevance and reputation.** Replication plays an essential self-correcting role in the systematic accumulation of knowledge by ensuring original findings are not due to error or chance. It also provides a deterrent effect: if one must share data and workflow, and if there is reasonable certainty one's research will be replicated by others, then QRPs will be less likely ... **Criminology should encourage a culture that supports open science, funds more replications, carries out and publishes more replications, and advocates for policy and practice based on reproducible findings.**"

# Barriers to Open Data

**Table 1.** Conflict between interests of researchers and advancement of science.

argument	counter-argument
1. Lack of time to curate data.	Unless adequately curated, data will over time become unusable, including by the original researcher [7].
2. Personal investment—reluctance to give data to freeloaders.	Reuse of data increases its value and the researcher benefits from additional citations [8]. There is also an ethical case for maximizing use of data obtained via public funding [9].
3. Concerns about being scooped before the analysis is complete.	This is a common concern though there are few attested cases. A time-limited period of privileged use by the study team can be specified to avoid scooping [10].
4. Fear of errors being found in the data.	Culture change is needed to recognize errors are inevitable in any large dataset [11] and should not be a reason for reputational damage. Data-sharing allows errors to be found and corrected.

Bishop, D. V. M. (2016). Open research practices: Unintended consequences and suggestions for averting them. (Commentary on the peer reviewers' Openness initiative). Royal Society Open Science, 3, 1–6. <https://doi.org/10.1098/rsos.160109>

## Practical Tips for Ethical Data Sharing

**Michelle N. Meyer**

Geisinger Health System, Danville, Pennsylvania

Advances in Methods and  
Practices in Psychological Science  
2018, Vol. 1(1) 131–144  
© The Author(s) 2018  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/2515245917747656  
www.psychologicalscience.org/AMPPS  


"One important tool in the reproducibility tool kit is data sharing, which enables reanalysis, replication, and well-powered consortium science. Historically, **IRBs and many researchers have prioritized data secrecy over data sharing.** Participants do often have privacy interests that are important to consider. Consequently, **[participants] should be asked for their permission to share their data, and care should be taken in deciding how and with whom their data are shared.** But it is past time for the research community to realize that participants typically also expect that the data they contribute will be used to advance scientific truth, not merely to make scientific claims that cannot be verified."

# Where to share data?



Discipline-specific

Institutional

General purpose

UK Data Service



Apollo

zenodo

**ALWAYS check legal and ethical restrictions first**

# Follow the FAIR principles

---

Findable – easily discoverable

---

Accessible – no restriction to access

---

Interoperable – doesn't rely on one type of software

---

Reusable – clearly licensed



## Part 5: Data management plans (DMP)



# What is a DMP?

## Summary document (2-3 pages)

- Outlines data management during and after research project
- Optional for graduate students (but recommended!)

## Purpose and benefits

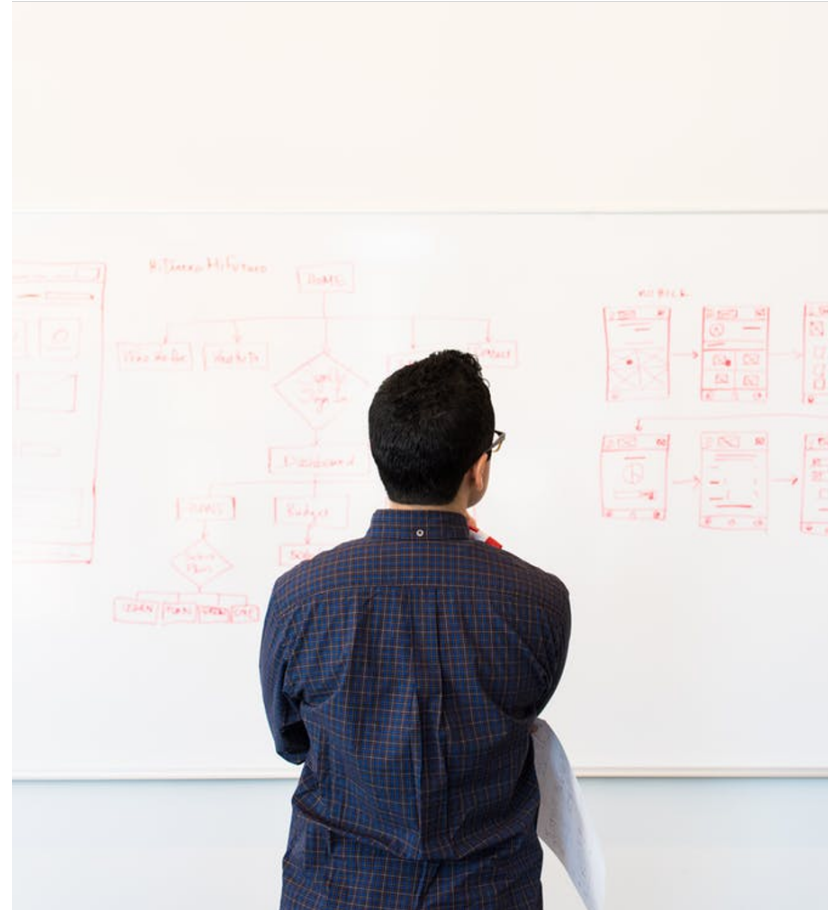
- Develop strategies for data storage, retention, and sharing
- Anticipate potential legal and ethical issues
- Estimate project costs for grant applications

## Funder requirements

- ... but also improves integrity and efficiency of research

# What does a DMP contain?

1. Sources of data
2. Quality assurances
3. Security and backup strategies
4. Data sharing plans
5. Ethical and legal restrictions
6. Data responsibility



[Home](#)
[About](#)
[Future plans](#)
[Help](#)
[Change language](#)

# Welcome.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3).

## Screenshot on how to use DMPonline

## Sign in



[Forgot your password?](#)

☐ Remember me

[Sign in](#)

[Or, sign in with your institutional credentials \(UK users only\)](#)

## Create account

New to DMPonline? Create an account today.

[Contact us](#) | [Terms of use](#)

© 2004 - 2017 Digital Curation Centre (DCC)

Digital Curation Centre DMP guidance: <https://www.dcc.ac.uk/dmps>



## Part 6: Further resources



# Cambridge Research Data services



[www.data.cam.ac.uk](http://www.data.cam.ac.uk)

# Research Integrity and Ethics

The University of Cambridge is committed to achieving excellence in research and scholarship. The pursuit of excellent research and the fulfilment of our responsibilities to participants in research, research users and the wider community require the maintenance of the highest standards of integrity and ethics. These pages provide guidance on and links to University policies addressing issues of research integrity, research ethics and good research practice.



Research Integrity



Research Ethics



Research Misconduct



Training



<https://www.research-integrity.admin.cam.ac.uk>

# Robust Behavioural Science Course

The screenshot shows the Department of Psychology website. The header includes a navigation bar with links: Home, About Us, People, Study, Research, Postdocs, Work with us, Services, and Intranet. Below the header, the page title is "Conducting Robust Behavioural Science: Syllabus (Raven)". The main content area features a video player for "Robust Behavioural Science: Introduction (Week 0)" by Dr Amy Orben, Michaelmas Term 2020. The video player has a play button and a title "Robust Behavioural Science" with the presenter's name and term. To the right of the video player is a sidebar with links for "Prospective Postgraduates", "Current Postgraduates" (with sub-links for PhD students and MPhil Biological Science), "General Information", "Resources for Postgraduate Students", "Courses and Seminars", "Ethics Committee", "Postgraduate wellbeing", "Research Councils", and "ESRC DTP".

Department of Psychology

Home About Us People Study Research Postdocs Work with us Services Intranet

Conducting Robust Behavioural Science: Syllabus (Raven)

Dr Amy Orben  
Michaelmas Term 2020  
General Introduction video:

Robust Behavioural Science: Introduction (Week 0)

**Robust Behavioural Science**  
Dr Amy Orben  
Michaelmas Term 2020  
Overview

Powered by Panopto

**Prospective Postgraduates**

**Current Postgraduates**

- > PhD students
- > MPhil Biological Science

**General Information**

- Resources for Postgraduate Students
- Courses and Seminars
- Ethics Committee
- Postgraduate wellbeing

**Research Councils**

- ESRC DTP

<https://www.psychol.cam.ac.uk/study/grads/grads/conducting-robust-behavioural-science>



# Ask a Data Champion



Clare Trowell



**OSC**

Office of Scholarly Communication



**Research  
Data**



**Cambridge  
Data Champions**





# Discussion and questions



# Points for Discussion



What strategies do you currently use to store and back-up your data?



How are your digital and physical files currently organised?



What types of personal or sensitive data are you handling in your research?



What are the benefits and challenges of data sharing in your research area?